paper_draft

DM

Title

- the name of the organism studied

- the particular aspect or system studied

- the variable(s) manipulated.

Abstract

passive voice

(typically 150-300 words)

- A brief introduction to the topic that you're investigating.

- Explanation of why the topic is important in your field/s.

- Statement about what the gap is in the research.

- Your research question/s / aim/s.

- An indication of your research methods and approach.

- Your key message.

- A summary of your key findings.

- An explanation of why your findings and key message contribute to the field/s

Introduction

Neurodevelopment disorders are a diverse and heterogeneous group of conditions impacting the development of the nervous system, brain function, physical development, emotional development and learning ability. A recently identified condition known as Neurodevelopmental Disorder with Microcephaly, Arthrogryposis, and Structural Brain Anomalies (NEDMABA) (MIM:618622) has been described in children with biallelic loss-of-function variants in *SMPD4* (Magini et al. 2019).

Sphingomyelinases such as *SMPD4* hydrolyze sphingomyelin into ceramide and phosphorylcholine. *SMPD4* specifically encodes one of 4 neutral sphinogmyelinases, nSMase3 (MIM: 610457).

So far, only approximately 30 cases have been reported across over 12 families. Clinical phenotype data of these cases is largely heterogeneous with severe neurological complications and early demise a key feature. This makes the identification, diagnosis and ongoing management of cases a challenge for patients, their families and medical practitioners.

Ravenscroft et al. (2021) performed a study over 190 probands with a diagnosis of arthrogryposis multiplex congenita, distal arthrogryposis, fetal akinesia deformation sequence or multiple pterygium syndrome. This study identified a novel missense variation in SMPD4 which at the time of whole exome sequencing was not well described in the literature. The impacted family from Melbourne exhibited features involving arthrogryposis multiplex congenita, complex brain malformations, small for gestation age and hypoplasia of the corpus callosum. In two of the three related cases were additional features of microcephaly, congenital encephalopathy, cerebellar malformation and hypoplasia and hypomyelination.

Monies et al. (2019) described two families with different homozygous truncating variants in SMPD4 as a syndrome of skeletal dysplasia with cerebella atrophy. The phenotypes described involved bilateral clenched hands, talipes, IUGR, partial absence of corpus callosum with family history of three neonatal deaths with similar features.

A detailed study by Magini et al. (2019) involved 12 unrelated familes with 32 individuals (21 with detailed clinical information). Hallmark presentations were of microcephaly, simplified gyration, hypomyelination, thin corpus callosum, mild cerebellar hypoplasia, brainstem hypoplasia, congenital arthrogryposis, diabetes mellitus, heart disease, severe encephalopathy and respiratory problems often leading to early demise. Despite this being the largest cohort studied, the clinical features and survival times among participants varied greatly. Three missense changes were noted in the study with affected children in these families often showing a milder presentation suggestive of possible residual function. In these cases individuals were able to develop independent motor skills, have mild intellectual disability and arthrogryposis without evidence of simplified gyral patterns on brain MRI. Other patients with truncating variants are shown to have more severe presentations, while a range of additional significant phenotypes were reported involving dysmorphic facial features, seizure, vocal cord paralysis and hearing impairment.

A recent case study from China is described in Ji et al. (2022) involving a girl presenting in infancy with intrauterine growth restriction, microcephaly, postnatal developmental delay, arthrogryposis, hypertonicity, seizure, and hypomyelination on brain magnetic resonance imaging. Ji et al. (2022) argues the actual prevalence of this condition, based on the gene carrier rate is understated and may be attributable to typical symptoms of NEDMABA being non-specific, providing a diagnostic challenge in a clinical setting.

A recent study by Bijarnia-Mahay et al. (2022) presents the case of a 22-month old girl presenting with the typical phenotype of neurodevelopmental delay, prenatal onset growth failure, arthrogryposis, microcephaly and brain anomalies including severe hypomyelination, simplified gyral pattern and hypoplasia of corpus callosum and brainstem. Notably, there is also additional non-typical clinical findings of nystagmus and visual impairment secondary to macular dystrophy and retinal pigment epithelial stippling at posterior pole.

Ji et al. (2022) reports parallels with two cases showing the same homozygous null variant. The individual reported in Monies et al. (2019) presented with distinct symptoms of brain atrophy and skeletal dysplasia whereas the case in Magini et al. (2019) with the same variant exhibited more typical clinical features.

Further work to collate and analyse the current body of research is challenging. While Magini et al. (2019) cataloged a detailed clinical phenotype data set as a supplimentary to their study, these data are not suitable for statisical analysis. Many of the clinical features are represented as free text descriptions and a variety of non-standard terminology is applied between cases, making machine interpretation of the data difficult. Other studies (Ji et al. 2022; Bijarnia-Mahay et al. 2022) present only written case reports with some tabulated summaries. While larger studies (Ravenscroft et al. 2021; Monies et al. 2019) offer data genotype analysis, there is less detail included on the clinical phenotype of indivuduals in the cohort.

Given the severe nature of this condition, often leading to early demise, a gap exists in understanding the expected survival time for individuals with bi-allelic null-variants in SMPD4. While a connection has been highlighted between some missense variants and a milder presentation with longer survival, this hypothesis has not been analysed further. Furthermore, the associations between diverse features of NEDMADA have not been analysed in detail, in-part due to the sparse and inconsistent collection of clinical and case report data for this new condition.

An open research question exists around the clinical pathway and survival time of children exhibiting bi-allelic loss-of-function variants in SMPD4. With current research highlighting a diverse and heterogeneous phenotype, the relationship and correlations between these diverse features is not well understood.

This research has three key aims. First, to collate and transform the variety of early case reports and studies on clinical phenotype data for this novel variant into an analysis-ready *tidy* dataset (Wickham 2014). Secondly, to statistically quantify the expected survival time of children with NEDMABA based on the current case reports. Finally, detailed analysis will

be conducted to further understand the associations between both patients and their clinical features.

While current research in this area has highlighted a diverse range of presentations, this will be the first in-depth analysis of early studies of this novel variant, which aims to produce statistical findings to better describe and understand this condition. This research is significant as it will assist clinicians understand the typical and non-typical presentations of this challenging new condition. In addition, genetic counselling of families with affected children will benefit from enhanced analysis on outcomes of existing reported cases.

Methods

SMPD4 Data Package

The data contained in Magini et al. (2019) "Summary of SMPD4-Related Clinical Phenotype" is an excel spreadsheet tabulating each of the 21 individuals from the study with clinical details recorded. The file has 21 columns (one for each individual), and 64 rows (one for each phenotype or clinical remark).

The data were read into R (R Core Team 2022) without any changes, so as to preserve the reproducibility of the data transformation steps.

The data were transposed to ensure it could be presented as *tidy* formatted data (Wickham 2014). This requires:

1. Each variable (clinical phenotype) forms a column.
2. Each observation (individual) forms a row.
3. Each type of observational unit forms a table (every cell has just one item)

In many cases, key clinical information was entered as free-text descriptions, which rendered any attempt of meaningful analysis impractical (Table 1). In these cases, the text was tokenised by separating the list of clinical observations at each comma and forming a binary indicator column noting its presence '1' or absence '0' (Table 2).

Table 1: Example of non-tidy free-text descriptions of features

	Family 1- Individual 1	Family 1- Individual 4
Facial dysmorphisms	short palpebral fissures, large ears, simple helices, smooth philtrum, thin lips, bilateral simian creases	short palpebral fissures, receding forehead, thin upper lip

Table 2: Example of tidy formatted data where individuals are transposed into rows and text into binary indicators

id	short palpebral fissures	large ears	simple helices	receding forehead	...
Family 1- Individual 1	1	1	0	...	
Family 1- Individual 4	1	0	0	1	...

In the case where two variables were formed from phenotypes that are considered to be synonymous, these were merged into one indicator column to prevent duplication e.g. {bilateral_cleft_lips, bilateral_cleft_lip, cleft_lip_b_l}.

Some data type conversion and categorical level standardisation was performed to ensure the data were in consistent and appropriate data types. For example, 'Gender' was not consistently coded, and 'Birth Weight' was encoded as a text string rather than a more useful numeric format. (Table 3).

Table 3: Example of inconsistent coding or sub-optimal data types

Gender	Birth Weight
male	2175 grams (- 2.5 SD)
female	2045 g (-3 SD)
Female	2300 gram (-2 SD)
Female fetus	n.a.

The final dataset consisted of 21 observations (one per individual) and 152 variables (one per clinical feature).

This format was preserved and other case studies identified in the literature (Ravenscroft et al. 2021; Monies et al. 2019; Bijarnia-Mahay et al. 2022; Ji et al. 2022) were manually entered to conform to this template to allow the data to be combined for further analysis.

These data sets were packaged into an R Package (Marchiori 2022) in order to allow for reproducibility and sharing. This can be downloaded and installed from github at https://github.com/deanmarchiori/SMPD4.

Multiple Correspondence Analysis

Cluster Analysis

Survival Analysis

- Explain why each procedure was done

- Experimental procedures and results are narrated in the past tense

- Mathematical equations and statistical tests are considered mathematical methods and should be described in this section along with the actual experimental work.

Results

The function of this section is to summarize general trends in the data without comment, bias, or interpretation. The results of statistical tests applied to your data are reported in this section although conclusions about your original hypotheses are saved for the Discussion section.

Discussion

- the relationship between the results and the original hypothesis, i.e., whether they support the hypothesis, or cause it to be rejected or modified

- an integration of your results with those of previous studies in order to arrive at explanations for the observed phenomena

- possible explanations for unexpected results and observations, phrased as hypotheses that can be tested by realistic experimental procedures, which you should describe

- End the Discussion with a summary of the principal points you want the reader to remember

Bijarnia-Mahay et al. (2022)

References

Bijarnia-Mahay, Sunita, Puneeth H Somashekar, Parneet Kaur, Samarth Kulshrestha, Vedam L Ramprasad, Sakthivel Murugan, Seema Sud, and Anju Shukla. 2022. "Growth and Neurodevelopmental Disorder with Arthrogryposis, Microcephaly and Structural Brain Anomalies Caused by Bi-Allelic Partial Deletion of Smpd4 Gene." *Journal of Human Genetics* 67 (3): 133–36.

Ji, Weigang, Xiangtian Kong, Honggang Yin, Jian Xu, and Xueqian Wang. 2022. "Case Report: Novel Biallelic Null Variants of Smpd4 Confirm Its Involvement in Neurodevelopmental Disorder with Microcephaly, Arthrogryposis, and Structural Brain Anomalies." *Frontiers in Genetics* 13.

Magini, Pamela, Daphne J Smits, Laura Vandervore, Rachel Schot, Marta Columbaro, Esmee Kasteleijn, Mees van der Ent, et al. 2019. "Loss of Smpd4 Causes a Developmental Disorder Characterized by Microcephaly and Congenital Arthrogryposis." *The American Journal of Human Genetics* 105 (4): 689–705.

Marchiori, Dean. 2022. *Smpd4: Smpd4-Related Clinical Phenotype Data.* https://deanmarchiori.github.io/SMPD4.

Monies, Dorota, Mohammed Abouelhoda, Mirna Assoum, Nabil Moghrabi, Rafiullah Rafiullah, Naif Almontashiri, Mohammed Alowain, et al. 2019. "Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population." *The American Journal of Human Genetics* 104 (6): 1182–1201.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ravenscroft, Gina, Joshua S Clayton, Fathimath Faiz, Padma Sivadorai, Di Milnes, Rob Cincotta, Phillip Moon, et al. 2021. "Neurogenetic Fetal Akinesia and Arthrogryposis: Genetics, Expanding Genotype-Phenotypes and Functional Genomics." *Journal of Medical Genetics* 58 (9): 609–18.

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10): 1–23. https://doi.org/10.18637/jss.v059.i10.